Editorial

Composite Versus Individual Measures of Disease Activity in Rheumatoid Arthritis


The measurement of disease activity in rheumatoid arthritis (RA) reflects the complicated nature of the disease: elements include the number of tender and swollen joints, pain, function, inflammation, and an overall assessment by both the individual and the physician. Historically, RA clinical trials assessed and reported a variable number of these components, making it difficult to compare therapies; this approach also allowed discretion regarding which measures to report. However, in the last 15 years, it has become standard to use composite indices such as the American College of Rheumatology (ACR) core set of disease activity measures (i.e., ACR20, ACR50, ACR70)[1] and the Disease Activity Score (DAS)[2] in the assessment of therapeutic efficacy in clinical trials of RA. As a consequence, trials use a common standard outcome (e.g., ACR20 or DAS), improving our ability to differentiate effective treatments from ineffective ones and to compare results across trials.

Do such composite measures of disease activity offer any advantages over individual measures with respect to sensitivity to change?

An ideal outcome measure has high sensitivity to change, which means that it differentiates between improvement and non-improvement with a high degree of precision, allowing effective treatment to be distinguished from placebo with a minimal sample size in a clinical trial. Just as hypertension alone has prognostic importance in cardiovascular disease, so too tender joint count reflects disease activity in RA. As a combined measure, the Framingham Risk Score[3], which includes age, sex, total cholesterol, high density lipoprotein cholesterol, smoking, and systolic blood pressure, is a still better predictor of cardiovascular disease versus a single measure. So, too, the core set measures in the ACR20 and the DAS better distinguish RA-related disease activity versus individual measures.

What are the properties of a composite measure that allow for this improved ability to distinguish treatment efficacy between trial arms?

Certain core set measures, such as swollen joint count, are less than optimal because they are not particularly sensitive to change[4-6], and a measure of functional limitation [Health Assessment Questionnaire (HAQ)] may have poor sensitivity to change in those with greater disease duration[7]. Despite disadvantages to these individual measures of disease activity, the composite measures that include them are, in fact, quite sensitive to change, allowing us to distinguish efficacious from non-efficacious treatments.

A composite index is more sensitive to change because its precision is usually better than that of its individual components. Because of its covariance structure, a composite measure generally yields less variability or noise than individual measures[8]. This improved precision can be achieved by combining measures that correlate with one another modestly[9]. Items that are too highly correlated fail to add additional information to a composite index[10]. In RA, all individual patient-derived core set measures (e.g., pain, functional limitation, global assessment) correlate highly with one another, and although most of these are sensitive to change, their combination into a patient composite measure may have poorer sensitivity to change than other composites that combine more diverse core set measures even if the individual measures being combined do not perform especially well on their own. In addition to their sensitivity to change, composite measures reflect the complex reality of RA disease activity, with contributions from blood markers of inflammation, pain, joint swelling and effects of joint pain and other contributors to a person's function. A valid measure of disease activity should include elements of all of these measures, irrespective of its performance.

Pincus, *et al*[11], in this issue of *The Journal*, evaluate the relative efficiencies of individual core set measures in 4

See Relative efficiencies of joint counts are not greater than physician/assessor global estimates and patient questionnaire measures to distinguish adalimumab from control treatments in RA RCT, *page 201*

clinical trials of adalimumab. If an outcome measure has a higher relative efficiency versus the tender joint count, then the trial sample size required to distinguish active treatment from placebo when using this measure will be smaller than that using tender joint count. The authors concluded that no single core set measure was substantially more efficient than any other measure. In fact, despite the trials being of the same agent, each trial showed a different core set measure as having the best sensitivity to change, with C-reactive protein, assessor global estimate, and patient global estimate performing the best in separate trials (see Table 2 in Pincus, *et al*[11]). For the most part, the differences in relative efficiencies of the core set measures were small and not statistically different from one another. As others have suggested[1], assessor global estimate performs as well as or better (and in some cases substantially better) than the rest of the core set measures in most cases.

What can we learn from these findings?

First, even though all 4 trials evaluated the same drug, the variability from trial to trial as to which core set measure was most sensitive to change may reflect differences in the patient populations enrolled and the concomitant therapies allowed. However, the more likely explanation is that there is chance variation from study to study in the performance of these measures. Since composite measures limit the variation or noise of measurement, they also eliminate some of this trial-specific chance variation. For example, the composite ACR20 response rates for those assigned to 40 mg of adalimumab (every other week in 3 trials[12-14], weekly in one trial[15]) were all statistically significantly and substantially better versus placebo. While Pincus, *et al* did not compare the efficiency of individual measures to composite measures, others have shown in such comparisons that the relative efficiencies of composite measures are almost always greater than those of each individual measure[5,16,17]. Indeed, Pincus, *et al* have shown impressive efficiency for composite measures in a leflunomide trial[18].

Pincus, *et al* note that the assessor global estimate performed well overall. This is not surprising since this measure is itself a type of composite; most physicians assign a global disease activity score taking into account clinical data at hand: patient-reported history of pain and functional limitations, physical examination involving assessment of joint tenderness and swelling, and laboratory variables. However, that the assessor global estimate did not perform substantially better than other core set measures in each of the trials points to the need for an RA trial composite index that includes measures in addition to this global assessment. RA composite indices contain multiple domains not all expressed by the assessor global estimate alone.

Sensitivity to change is of great importance in the design and conduct of RA clinical trials since it is the prime determinant of sample size required to distinguish active treatment from placebo (or to compare 2 active treatments). As RA trials move towards comparing treatment regimens that may not offer vast differences in efficacy, outcome measures must be sensitive enough to detect smaller differences in treatments. Less sensitive outcome measures would require unrealistic sample sizes. The need to distinguish smaller treatment differences has already led to the development of the ACR hybrid[19], which utilizes the established ACR core set measures and improves upon the performance characteristics of the ACR composite index by proposing a new way to combine the individual measures to maximize sensitivity to change.

The findings of Pincus, *et al*[11] provide evidence that investigators must be cautious when considering simplifying clinical trial outcome assessments by relying solely on one particular measure. Such a strategy, while possibly reducing participant burden, would increase the number of participants needed in trials. On the other hand, the burden of obtaining the data required to compute composite measures may be problematic in clinical practice. As Pincus, *et al* point out, "a formal quantitative joint count, which is not performed by most rheumatologists at most visits[20], may not be necessary to monitor patients quantitatively in a busy clinical care setting." Of the core set, those measures that are patient-derived have been shown to predict longterm outcomes of relevance to clinical practice, such as disability and death[21-26]. Although the current article[11] does not address the utility of individual core set measures in the clinical setting, one could infer that a patient or physician global assessment may be sufficient in clinical practice to track an individual's disease activity over time. In such a setting, sensitivity to change is less important than monitoring disease activity, and one must consider the practical burden of obtaining each core set measure at each visit. To that end, Pincus, *et al* provide indirect support for the physician global estimate as a key disease activity measure in the clinical setting, bearing in mind that this global estimate is usually informed by other components of the core set and, as noted above, is a type of composite measure. Even so, collecting data on all core set measures in the clinic setting would ensure that the rheumatology community continues to use a common metric that helps evaluate the longterm "real-world" effectiveness of therapies whose efficacy has been demonstrated only in the selected patients eligible for short-term clinical trials[27]. This is particularly pertinent to clinical settings that are also research clinics.

In conclusion, while Pincus, *et al*'s study provides useful information about the comparative efficiency of individual core set measures in RA, it is not clear how relevant this information is and for what setting. Clinical trials and clinical practice have different purposes and endpoints. For RA trials, it matters little whether specific individual measures have more or less sensitivity to change than others: RA trials assess change using composite indices that depend both

on sensitivity to change of individual measures and on the correlation of these measures with one another. For clinical practice, where it may be burdensome to measure all core set measures including joint counts at each visit, the excellent sensitivity to change and composite nature of global assessments suggest they are critical elements in the evaluation of disease activity. However, even in the clinic setting, information on all core set measures provides an opportunity to evaluate the longterm real-world performance of newer therapies.

TUHINA NEOGI, MD, FRCPC;
DAVID T. FELSON, MD, MPH,
Clinical Epidemiology Research and Training Unit,
Boston University Medical Center,
Boston, Massachusetts, USA

Supported by NIH AR47785. Address reprint requests to Dr. T. Neogi, Boston University School of Medicine, 650 Albany St., Clinical Epidemiology Unit, Suite X-200, Boston, MA 02118, USA.

REFERENCES

1. Felson DT, Anderson JJ, Boers M, et al. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. Arthritis Rheum 1993;36:729-40.
2. van der Heijde DM, van 't Hof M, van Riel PL, van de Putte LB. Development of a disease activity score based on judgment in clinical practice by rheumatologists. J Rheumatol 1993;20:579-81.
3. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. Circulation 1998;97:1837-47.
4. Anderson JJ, Bolognese JA, Felson DT. Comparison of rheumatoid arthritis clinical trial outcome measures: a simulation study. Arthritis Rheum 2003;48:3031-8.
5. Anderson JJ, Chernoff MC. Sensitivity to change of rheumatoid arthritis clinical trial outcome measures. J Rheumatol 1993;20:535-7.
6. Tilley BC, Pillemer SR, Heyse SP, Li S, Clegg DO, Alarcon GS. Global statistical tests for comparing multiple outcomes in rheumatoid arthritis trials. MIRA Trial Group. Arthritis Rheum 1999;42:1879-88.
7. Aletaha D, Strand V, Smolen JS, Ward MM. Treatment-related improvement in physical function varies with duration of rheumatoid arthritis. A pooled analysis of clinical trial results. Ann Rheum Dis 2007; Jul 20 Epub ahead of print.
8. Nunnally JC. Psychometric theory. 2nd ed. New York: McGraw-Hill; 1978.
9. Anderson JJ. Mean changes versus dichotomous definitions of improvement. Stat Methods Med Res 2007;16:7-12.
10. Streiner DL, Norman GR. Selecting the items. In: Health measurement scales: A practical guide to their development and use. New York: Oxford University Press; 2003.
11. Pincus T, Amara I, Segurado OG, Bergman M, Koch GG. Relative efficiencies of joint counts are not greater than physician/assessor global estimates and patient questionnaire measures to distinguish adalimumab from control treatments in rheumatoid arthritis clinical trials. J Rheumatol 2008;35:201-5.
12. Furst DE, Schiff MH, Fleischmann RM, et al. Adalimumab, a fully human anti-tumor necrosis factor-alpha monoclonal antibody, and concomitant standard antirheumatic therapy for the treatment of rheumatoid arthritis: results of STAR (Safety Trial of Adalimumab in Rheumatoid Arthritis). J Rheumatol 2003;30:2563-71.
13. Keystone EC, Kavanaugh AF, Sharp JT, et al. Radiographic, clinical, and functional outcomes of treatment with adalimumab (a human anti-tumor necrosis factor monoclonal antibody) in patients with active rheumatoid arthritis receiving concomitant methotrexate therapy: a randomized, placebo-controlled, 52-week trial. Arthritis Rheum 2004;50:1400-11.
14. Weinblatt ME, Keystone EC, Furst DE, et al. Adalimumab, a fully human anti-tumor necrosis factor alpha monoclonal antibody, for the treatment of rheumatoid arthritis in patients taking concomitant methotrexate: the ARMADA trial. Arthritis Rheum 2003;48:35-45.
15. van de Putte LB, Rau R, Breedveld FC, et al. Efficacy and safety of the fully human anti-tumour necrosis factor alpha monoclonal antibody adalimumab (D2E7) in DMARD refractory patients with rheumatoid arthritis: a 12 week, phase II study. Ann Rheum Dis 2003;62:1168-77.
16. Felson DT. A proposed revision to the ACR20: The hybrid measure of the American College of Rheumatology response. Arthritis Rheum 2007;57:193-202.
17. Felson DT, Anderson JJ, Boers M, et al. American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum 1995;38:727-35.
18. Pincus T, Strand V, Koch G, et al. An index of the three core data set patient questionnaire measures distinguishes efficacy of active treatment from that of placebo as effectively as the American College of Rheumatology 20% response criteria (ACR20) or the Disease Activity Score (DAS) in a rheumatoid arthritis clinical trial. Arthritis Rheum 2003;48:625-30.
19. American College of Rheumatology Committee to Reevaluate Improvement Criteria. A proposed revision to the ACR20: the hybrid measure of American College of Rheumatology response. Arthritis Rheum 2007;57:193-202.
20. Pincus T, Segurado OG. Most visits of most patients with rheumatoid arthritis to most rheumatologists do not include a formal quantitative joint count. Ann Rheum Dis 2006;65:820-2.
21. Callahan LF, Pincus T, Huston JW 3rd, Brooks RH, Nance EP Jr, Kaye JJ. Measures of activity and damage in rheumatoid arthritis: depiction of changes and prediction of mortality over five years. Arthritis Care Res 1997;10:381-94.
22. Pincus T. The DAS is the most specific measure, but a patient questionnaire is the most informative measure to assess rheumatoid arthritis. J Rheumatol 2006;33:834-7.
23. Pincus T, Brooks RH, Callahan LF. Prediction of long-term mortality in patients with rheumatoid arthritis according to simple questionnaire and joint count measures. Ann Intern Med 1994;120:26-34.
24. Pincus T, Callahan LF, Sale WG, Brooks AL, Payne LE, Vaughn WK. Severe functional declines, work disability, and increased mortality in seventy-five rheumatoid arthritis patients studied over nine years. Arthritis Rheum 1984;27:864-72.
25. Soderlin MK, Nieminen P, Hakala M. Functional status predicts mortality in a community based rheumatoid arthritis population. J Rheumatol 1998;25:1895-9.
26. Wolfe F, Zwillich SH. The long-term outcomes of rheumatoid arthritis: a 23-year prospective, longitudinal study of total joint replacement and its predictors in 1,600 patients with rheumatoid arthritis. Arthritis Rheum 1998;41:1072-82.
27. Zink A, Strangfeld A, Schneider M, et al. Effectiveness of tumor necrosis factor inhibitors in rheumatoid arthritis in an observational cohort study: comparison of patients according to their eligibility for major randomized clinical trials. Arthritis Rheum 2006; 54:3399-407.